

音響教育を視野に入れた声道模型のための音源生成システムの開発*

☆前田 祐貴, 荒井 隆行, 安 啓一 (上智大)

1 はじめに

これまでの音響教育現場では、音響学的現象などを教科書やコンピュータのシミュレーションを用いて説明することが中心であった。しかし、文理を問わず多種多様な背景を持つ学習者に理解を促すには、より直感的な教材開発が望まれている。そのような中、音声生成に関しては、声道模型を用いた音響教育についてその有効性が確認されている（例えば、[1,2]など）。声道模型の使用は、母音生成の仕組みや音源フィルタ理論などを直感的に説明するのに適しており、また、大学生や大学院生だけでなく、子どもから専門家までの様々な人々に対して役立つことが報告されている[1, 2].

ところで従来の声道模型の音源としては、電動式人工喉頭や笛式人工喉頭などが用いられている[2]。また、ホーンスピーカのドライバユニットを使うことで、任意の信号を音源として用いることができる。いずれの場合も、ブザー音のような喉頭原音を声道模型に入力すると、模型から母音のような音声出力される。

本研究では、発せられる「生の音声信号」から実時間で声道模型のための音源信号を得るためのシステムを試作する。その際、入力音声のピッチ及び実効値 (RMS) 振幅に追従させることを考えた。このシステムが実現されれば、講義などのデモンストレーションのみならず、博物館や科学館において来館者が実時間で声を発し、その発した声による「音源」を用いて、また別の母音を発することができるようになり、声道模型に関してより一層理解が深まることが期待される。

2 音源生成のためのアルゴリズム

本システムで用いたピッチ及び RMS 振幅推定のためのアルゴリズムを Fig. 1 に示す。

2.1 ピッチの推定

ピッチ推定法には、零交差点を用いる方法[3]

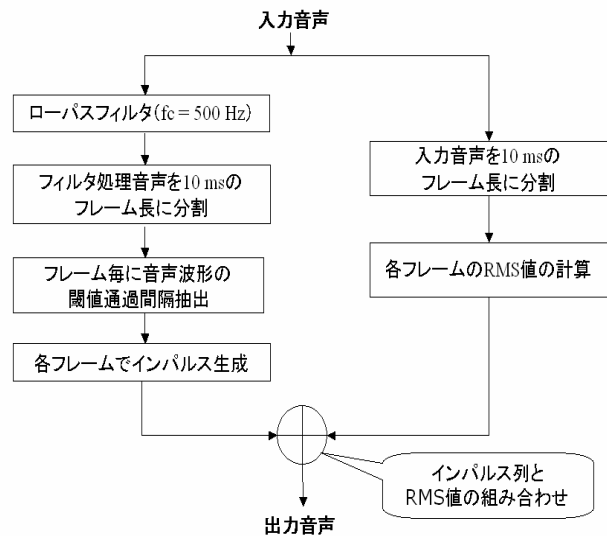


Fig. 1 アルゴリズムの概要

を応用し、音声波形がある閾値レベルを交差する「レベル交差点」にインパルスを立てる方法を用いた。まず、入力信号をローパスフィルタに通し、倍音成分を除去した。このローパス処理によって、高調波成分による影響が抑制される。ここで、子どもの音声の平均的な基本周波数を考慮し[4, 5], ローパスフィルタのカットオフ周波数を 500 Hz に設定した（標本化周波数は 16 kHz, タップ数は 1024 次）。

次に、ローパスフィルタに通した音声を 10 ms の長さのフレームに分割し、フレーム毎に振幅の閾値を設定した。ここで、男性の平均基本周波数が 125 Hz, 標準偏差が 20.5 Hz であることから[4-6], 入力信号の最低周波数を 100 Hz と仮定した。したがって、10 ms 長のフレーム内には最低でも 1 周期分の音声波形が含まれることになる。閾値は、そのフレーム内の最大振幅とフレーム内の RMS 振幅の中間値に設定した。閾値を交差した点において、振幅 1 のインパルスを立てることによって、入力信号のピッチに追従したインパルス列を生成した。

* Developing a Sound Source Generation System of the Vocal Tract Models for Education in Acoustics, by MAEDA, Yuki and ARAI, Takayuki, YASU Keiichi (Sophia University).

2.2 RMS 振幅の推定

入力信号の RMS 振幅を推定するために、まず入力信号をフレームに分割した。そしてフレーム毎に RMS 振幅を計算した。基本周期の影響をあまり受けずに RMS を計算するためにはある程度のフレーム長が必要である。しかし、長すぎると RMS 振幅の時間解像度が劣化する。そこで、RMS 振幅を計算するためのフレーム長は、ピッチ推定の際に仮定した最低基本周波数 100 Hz に対応する最長基本周期である 10 ms とした。

2.3 音声出力

入力信号波形の閾値交差点を用いて生成したインパルス列と、長さ 10 ms のフレーム毎に計算した RMS 値を掛け合わせた。以上の処理によって、入力信号のピッチ及び RMS 振幅に追従した出力信号を得る。入力信号波形・ローパスフィルタ処理波形・出力信号波形を Fig. 2 に示す。この際、出力信号はインパルス列のため、入力信号に比べ振幅が小さくなる。なお、本システムではサンプリング周波数を 16 kHz とした。

3 シミュレーション

Matlab を使用して、実際に音声を用いてシミュレーションを行った。

3.1 入力信号の種類

入力信号として、多言語音声コーパス[7]より母音からなる単語「いいえ」を用いた（男女各 6 人、計 12 発話）。その他、持続母音「あ」「い」「う」「え」「お」の 5 発話と、連続母音「あいうえお」の抑揚あり、なしの 2 発話を新たに録音し用いた。なお、録音場所は上智大学荒井研究室付属の防音室である。

3.2 評価

音声分析ソフト Praat (Ver. 5. 0. 20) [8]を用いて、入出力信号のピッチ及び RMS 振幅を抽出し、比較を行った。この際、ローパスフィルタによる出力信号の遅延を考慮した。

まずピッチについては、いずれの音声を入力しても、平均誤差は 1%以内にとどまった。一方、RMS 振幅については、出力信号はインパルス列であるため、入力信号よりも RMS 値が明らかに小さくなる。そのため、両者の比較を行う際、入出力信号の RMS 振幅において平均値を揃えた上で比較した。その結果、対数振幅の誤差は 1.8 dB

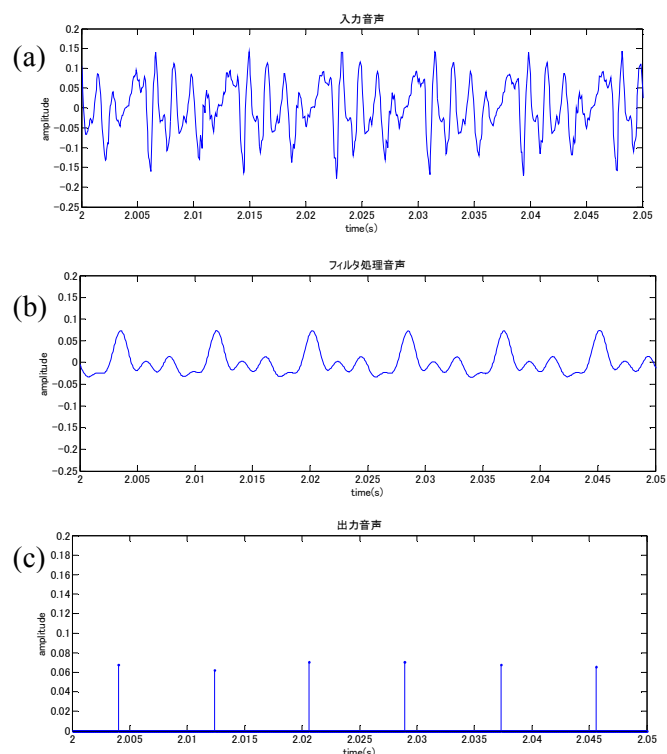


Fig. 2 入出力信号・フィルタ処理信号の波形：
(a) 入力 (b) フィルタ処理 (c) 出力

以内に収まった。聴覚印象としても、いずれの入力音声に対して出力音源のピッチ及び振幅は追従しているように聞こえた。

誤差の原因としては、ピッチに応じたインパルスの本数の変化が挙げられる。入力信号の基本周波数が低い場合と高い場合でインパルスの本数に差が生じてしまい、その結果 RMS 値にも差が生じる。よって、抑揚がある場合に、振幅に多少の誤差が生じたと考えられる。

3.3 声道模型への音源入力

実験で生成された各々の音源を、実際に声道模型に入力し、声道模型からの出力音声の確認を行った。その結果、それぞれの声道模型の形状 (/a/, /i/, /u/, /e/, /o/) のような音声出力されることが確認できた。

4 DSP を用いた実時間システム

4.1 概要

本実験で使用したシステムの DSP (digital signal processor) への実装を試みた。今回の実装では、都合により入力信号のピッチに追従したインパルス列を生成する際に、レベル交差ではなく零交差を用いた。そのため、出力信号には余計な

インパルスが含まれた. Matlab シミュレーションによる出力信号と, DSP による出力信号を Fig. 3 に示す. 図の入力信号は, 孤立発声の/a/である. DSP による出力信号には, 余分なインパルスが含まれてはいるが, 聴覚的には, 入力信号のピッチ及び振幅変化に追従しているように知覚できた. また, ホーンスピーカのドライバユニットからの出力を声道模型に接続することで, それぞれの声道形状に対応した母音出力を確認出来た.

ここで, Matlab シミュレーションによる出力信号と声道模型を用いた母音生成実験(実験 A)と, 肉声を入力とし, DSP による出力信号と声道模型を用いた母音生成実験(実験 B)を, 音声科学を学ぶ学生に実際に参加してもらい, その感想や意見を調査した. また, DSP を使用して声道模型から得られる音声を実際に, それぞれの模型形状に対応した母音として知覚されるのかを確認するために主観評価実験を行った. 実験参加者は, 上智大学理工学部情報理工学科の学生(3年次生)6人である. また, 実験は上智大学荒井研究室付属の防音室で行った.

4.2 アンケート調査

以下にアンケート内容と調査結果の一部を記す.

- ① 実験 A と実験 B において, どちらの方がより母音生成のしくみを理解しやすかったですか? また, その理由を記述してください.

⇒* 実験 B の方が, 実時間で入力音声に変化を与えることができ, 実際に自分自身で発した母音が, 声道模型からは異なる母音として聞こえた. その結果, 声道の形状によって母音に変化することを理解しやすかった.

* 実験 A のような予め処理した音声よりも, 自分の声を実時間で処理されることの方が説得力を感じられた.

- ② 実験 B では, 自分の発した母音とは異なる母音が声道模型から出力されましたが, どのように感じましたか?

⇒* どの母音を発しても, 声道模型の形状によって母音が決めるのは, 実際に体験してみると不思議な気分だった. 声道の形状によって, 母音の種類が決定されることを, 実験 B で深く実感できた.

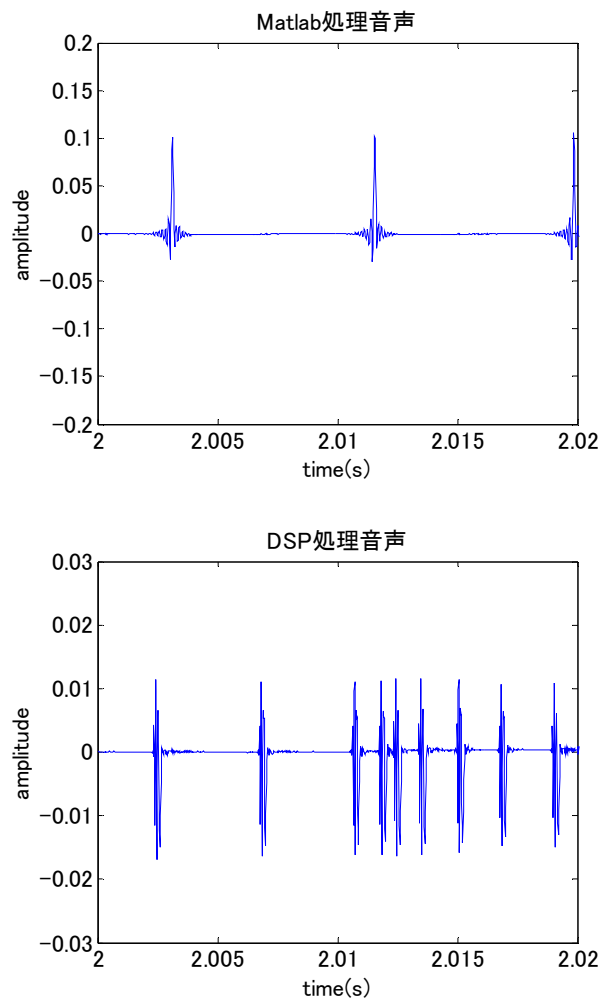


Fig. 3 処理音声波形 (Matlab 及び DSP 処理音声)

* 自分の発した母音とは異なる母音が出力されて, 誰かが自分の音声を利用して声を出しているような感じを受けた.

- ③ 実験 B では自分の発した声が高い(低い)と声道模型から生成される母音も高く(低く)なることを体験しました. 母音の高さは, 音源の高さによって決まることを知っていましたか?

⇒* 実験前までは知らなかった. いろんなパターンの高さの音声を入力し, 声道模型からの出力音声を聞くことで実感できました.

4.3 聴取実験

4.3.1 実験で使用した刺激及び手順

システムへの入力音声として, 持続母音/a/, /i/, /u/, /e/, /o/の5発話を使用し, それぞれを

DSP で処理した。DSP からの出力を声道模型（5 種類）に接続して得られる母音の録音を行った。録音音声は 25 通りである。持続時間は約 1 秒程度である。

それぞれの録音音声ランダムに並び替えてスピーカから再生し、聴こえた音声を書き取ってもらった。なお、録音・実験場所は上智大学荒井研究室付属の防音室である。

4.3.2 結果

聴取実験の回答結果を Table 1 に示す。Table 1 より、入力母音による回答結果の違いは見られず、声道形状/o/以外では高い正解率を得た。また、声道形状/o/の場合のみ正解率が低く、先行研究[1]と同様の結果であることが示された。

5 まとめ

本研究では、声道模型の音源として、入力信号のピッチ及び振幅に追従した音源生成を行うシステムを提案し評価を行った。まずシミュレーションによって、数種類の入力信号をシステムに通した際の入出力信号のピッチ及び振幅を比較した。その結果、ピッチの誤差はほとんどなく、入力信号のピッチを追従出来ていることが示された。振幅の比較では、入力信号のピッチが変化している際に多少の誤差が生じた。しかし、聴覚的には誤差は感じられなかった。また、入力母音の韻質による影響は見られなかった。

また、本システムで生成された音源を実際に声道模型に入力した結果、声道模型からそれぞれの声道形状に対応した母音が出力されることが確認できた。さらに、本システムの DSP への実装を試みた。レベル交差ではなく零交差による点にインパルスを立てたため、出力信号には余分なインパルスが含まれてしまったものの、実際にデモンストレーションの実験を行った結果、DSP を使用したシステムの方が、母音生成のしくみや声道フィルタ理論の理解が深まったという感想が大半を占めた。また、実時間で入力音声を変更させることが出来る点においても、DSP 上のシステムは評価が高かった。

現段階で DSP に実装されているプログラムを改良することで、余分なインパルスを含まず、入力信号のピッチ及び振幅に追従した出力信号を

Table.1 回答結果

声道模型	入力母音	回答				
		a	i	u	e	o
a	a	6	0	0	0	0
	i	6	0	0	0	0
	u	6	0	0	0	0
	e	6	0	0	0	0
	o	6	0	0	0	0
i	a	0	6	0	0	0
	i	0	6	0	0	0
	u	0	6	0	0	0
	e	0	4	0	2	0
	o	0	6	0	0	0
u	a	1	0	5	0	0
	i	0	0	6	0	0
	u	0	0	6	0	0
	e	0	0	6	0	0
	o	0	0	6	0	0
e	a	0	0	0	6	0
	i	0	0	0	6	0
	u	0	0	0	6	0
	e	0	0	0	6	0
	o	0	0	0	6	0
o	a	2	0	3	0	1
	i	0	0	6	0	0
	u	0	0	5	0	1
	e	0	0	5	0	1
	o	0	0	6	0	0

実時間で生成することが出来る。これによって、実時間で自分自身の音声音が音源に変換され、生成された音源を声道模型に入力すると、声道模型から母音が出力される。これにより音響学的現象を肌で感じる事ができ、母音生成の仕組みや音源フィルタ理論などの理解をさらに深めることが出来ると考えられる。また、博物館や科学館を始め、多くの人々に音響・音声学に興味を持ってもらうことが期待される。

参考文献

[1] T. Arai, *Journal of the Phonetic Society of Japan*, 5 (2), pp.31-38, 2001.
 [2] T. Arai, *Acoustical Science and Technology*, 28 (3), pp.190-201, 2007.
 [3] 村木健司他,通信技報, EA 93(119), pp.47-54, 1993.
 [4] R. D. Kent, A. D. Murray, *J. Acoust. Soc. Am.*, 72 (2), pp.353-365, 1982.
 [5] L. F. Halliday, et al., *J. Acoust. Soc. Am.*, 123(6), pp.4393-4402, 2008.
 [6] 中田和男, コロナ社, 1977.
 [7] 筑波大 多言語音声コーパス, 2001.
 [8] B. Paul and W. David, from <http://www.praat.org/>, 2009.